From Query to Opinion: The Impact of RAG Systems on User Attitude

NOAH MEISSNER, University of Regensburg, Germany

FELIX HORNBERGER, University of Regensburg, Germany

1 INTRODUCTION

Generative IR Systems, such as Chat GPT, Bing AI or Google Bard, offer an alternative means to address information needs [4, 9]. Such systems mark a significant progression in assisting users with complex search tasks, due to presenting a summarization of search results [8]. While these systems offer valuable support, the responses they generate are not immune to effects like hallucinations and system biases[8]. For that reason, Generative Information Retrieval systems employ Search Engine Results Pages (SERPs) through the use of retrieval-augmented generation (RAG)[3] techniques to substantiate their responses. Significantly, the biases present in Search Engine Results Pages (SERPs) can impact the results too[8]. To deepen our understanding of this phenomenon, our objective is to scrutinize the impact of biases on users' perspectives concerning contentious topics. Despite existing knowledge, there are two pivotal uncertainties regarding the effect of biases on user attitudes. Firstly, we aim to explore how the stances of sources influence users' attitudes as a consequence of their utilization in generative Information Retrieval (IR) Systems. Secondly, we seek to determine whether there is a noticeable variation in user interaction patterns with generative IR, contingent upon the opinions presented in the references. The objective of this research is to examine the influence of biases on user attitudes. In a user study, participants will be tasked with acquiring additional information on a controversial topic using a Generative Information Retrieval (IR) System. The topics will be chosen based on the minimal opinion bias of the participants. This study will measure both the quantitative responses of the users and the qualitative data derived from their interactions with the system.

2 STUDY-DESIGN

2.1 Study type

Randomized Controlled Trial employing an experimental design, encompassing three distinct conditions (pro, contra, neutral). Each condition corresponds to a unique opinion background, characterized by search results pertaining to one of three variably debated topics.

Authors' addresses: Noah Meißner, Information Science Faculty, University of Regensburg, Regensburg, Germany, Noah.Meissner@stud.uni-regensburg.de; Felix Hornberger, Information Science Faculty, University of Regensburg, Regensburg, Germany, Felix.Hornberger@stud.uni-regensburg.de.

2.2 Research Questions

- RQ1: How does the presentation of Generative IR answers based on opinion sources influence user attitudes
- RQ2: Is there a discernible difference in user interaction patterns with generative IR based on the reference opinions?

3 HYPOTHESES

3.0.1 Hypothesis related to RQ1.

- Hypothesis1 (H1): The stance expressed in sources of retrieval-augmented-generation Systems influences users' attitudes towards the debated topic.
  Rationale. In several Retrieval Systems the stances of references influence the users' attitude [2][8]. Generative IR System with the use of f retrieval-augmented generation (RAG) [3] use those Retrieval Systems too and we predict, this effect will transfer to the use in a Generative IR System,

4 METHOD

4.1 Materials

4.1.1 Data. In our research, we utilized the same data points as those employed in the study conducted by Markus Bink et al. [1], enabling a comparative analysis of their findings in Feature Snippets with our observations in Generative Information Retrieval Systems. We incorporated two publicly accessible datasets, each containing search results comprising URLs, titles, and snippets, across eight controversial topics.

The first dataset, curated by Draws et al.[5] using the Bing API, encompasses search results and associated annotations on topics including obesity as a disease, zoos, cell phone radiation safety, bottled water, and social networking sites. The stance annotations for this dataset were originally obtained through crowdsourcing methods by the primary researchers Bink et al. [1]. Enhanced this dataset by adding logics annotations to a subset of the search results.

The second dataset was compiled through web scraping or API techniques, employing various queries on two renowned search engines regarding subjects such as school uniforms, intellectual property rights, and atheism. Researchers annotated this dataset, assigning each search result two types of viewpoint labels: one indicating the stance and the other denoting the logics, which include categories like inspired, popular, moral, civic, economic, functional, and ecological perspectives.

We chose three out of the eight available topics to guarantee that

- our study is comparable to the Feature Snippet Study [1]
- viewpoints were equilibrated, meaning that the average stance value for the documents pertaining to each selected topic approximated zero;
- documents possessed adequate credibility, thereby eliminating a potential confounding variable;
- a diversity of logics was accessible for our research.

The topics selected in alignment with these criteria were:

- Is obesity a disease?
- Should intellectual property rights exist?
- Should students have to wear school uniforms?

4.1.2 Generative IR System. The dataset comprises four websites for each topic, containing content that is fed into a Generative Information Retrieval (IR) system. Our system is structured in such a way as to simulate a Retrieval-Augmented Generation (RAG) pipeline for the purpose of controlling the outcomes. This system autonomously determines which sources to use and the extent of their utilization. There are three potential stances that can be adopted: Pro, Contra (against), and Neutral. These stances will be selected randomly. The stances have an influence on the references, as the provided sources correlate with the chosen stances. In this phase of the study, we aim to ascertain whether variations in the presentation of biased sources affect user perceptions. The IR system was developed as a part of this study and is powered by the OpenAI API [1]. We also use various prompts to ensure that the message reaches the user with the respective condition[2].
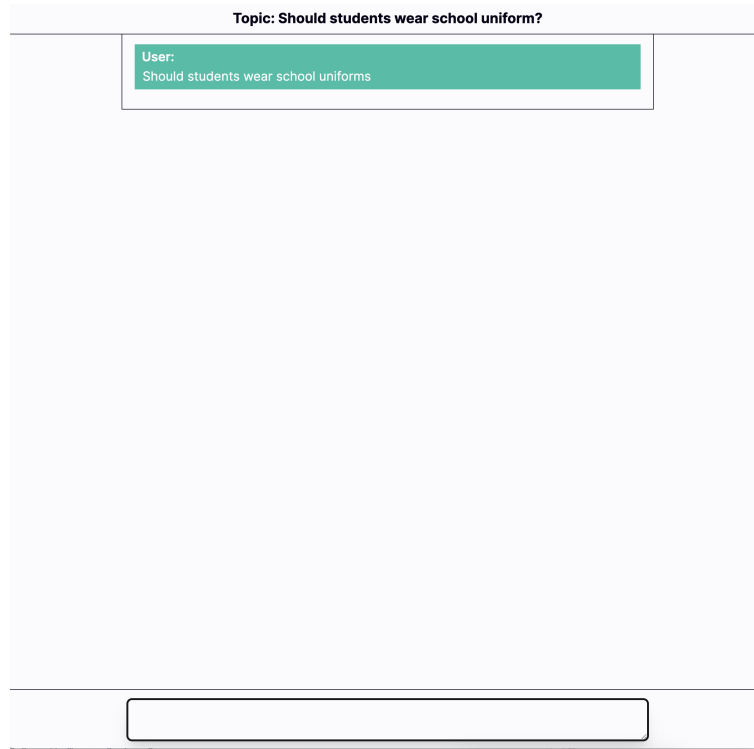

Topic: Should students wear school uniform?

User:
Should students wear school uniforms

Fig. 1. The user is starting a conversation and the system tries to answer the question with the material given

4.2 Variables

4.2.1 Independent Variables.

- Stance: in the Generative IR System

4.2.2 Dependent Variables.

- Attitude Change

---

[1] using GPT-3.5 Turbo (costing $0.0015 per 1000 tokens)
[2] more Information Fig: 2

4.3 Planned sample in pre-study

Participants for this study will be enlisted through university students. They will be provided with a compensation of 0.25 VP hours. The expected duration for each participant to complete the study ranges from 10 to 20 minutes.

4.4 Procedure

- (1) You will be initially provided with a brief introduction to the study. After this, you will complete the informed consent process.
- (2) You will be surveyed to evaluate your pre-existing attitudes on several topics. This step is crucial to ascertain your baseline familiarity with the topic. The system will specifically focus on topics where your stances fall within the range of -1 to 1.
- (3) You will be then introduced to the generative Information Retrieval (IR) System. You will engage in a dialogue with a Large Language Model, through which you access and explore information pertaining to the topic identified in step (2). Your conversation with the chatbot will be forwarded to third-party providers. Therefore, do not disclose any personal data in the chat and do not share any information that you would not discuss with other people. Only chat data will be shared with third party providers. Other data, such as demographics data, will not be shared outside the research team
- (4) When you have completed the task, you will be presented a short post-task questionnaire to ascertain if your perspective has changed and to gather further pertinent feedback and insights derived from your utilization of the system.

5 ANALYSIS PLAN

In order to conduct our study with rigor and statistical power, we have determined that a sample size of 67 participants is required. We performed a power analysis utilizing a one-sided Analysis of Variance (ANOVA) approach, employing the software G*Power [6, 7]. Our analysis considered an effect size (f) of 0.4, a significance threshold (alpha) of 0.05, and a desired statistical power of 0.8. The ANOVA will assess the impact of reference stances (pro, contra, neutral) and attitude changes as independent and dependent variables. Furthermore, we plan to perform additional exploratory analyses, which are not detailed in this document, with the aim of identifying potential trends within our dataset. These exploratory analyses will be thoroughly examined in the final script of our study. Through this analysis, we aim to address the research questions 2, focusing closely on their exploration.

REFERENCES

[1] Markus Bink. 2023. Preregistration: Investigating the influence of Featured Snippets on User Attitudes. https://osf.io/vghdx. (2023).

[2] Markus Bink, Sebastian Schwarz, Tim Draws, and David Elsweiler. 2023. Investigating the Influence of Featured Snippets on User Attitudes. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) *(CHIIR '23)*. Association for Computing Machinery, New York, NY, USA, 211–220. https://doi.org/10.1145/3576840.3578323

[3] Deng Cai, Yan Wang, Lemao Liu, and Shuming Shi. 2022. Recent advances in retrieval-augmented text generation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 3417–3419.

[4] Robert Capra and Jaime Arguello. 2023. How does AI chat change search behaviors? arXiv:2307.03826 [cs.HC]

[5] Tim Draws, Nava Tintarev, Ujwal Gadiraju, Alessandro Bozzon, and Benjamin Timmermans. 2021. This is not what we ordered: Exploring why biased search result rankings affect user attitudes on debated topics. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 295–305.

[6] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods* 41, 4 (2009), 1149–1160.

[7] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191.

[8] Ryen W White. 2023. Navigating Complex Search Tasks with AI Copilots. *arXiv preprint arXiv:2311.01235* (2023).

[9] Ruiyun Xu, Yue Feng, and Hailiang Chen. 2023. ChatGPT vs. Google: A Comparative Study of Search Performance and User Experience. arXiv:2307.01135 [cs.AI]

6 PROMPTS

System Prompt:
*You are a generative information retrieval system (IR system) that has the task of creating well-rounded opinions by carefully presenting arguments from different points of view. Your answers must endeavour to explain to the user the arguments given in the sources. To ensure the credibility and relevance of the information provided, you must use only the following pre-selected sources. **[SOURCES]** These sources have been selected for their reliability and comprehensiveness. When formulating your answer, explicitly refer to these sources to support your arguments. This not only increases the credibility of your answer, but also gives users the opportunity to study the original materials to gain further insight. When answering the user's question, make sure that you:*
- Present arguments from different perspectives to ensure a comprehensive and unbiased view.
- In your answer, refer directly to the sources provided and cite them appropriately to enable the user to check them.
- Maintain your position given in the sources
*The following is the chat history: **[Chat History]***
*This is the prompt of the user: **[Prompt of the User]***
Please proceed with answering the question as described above, adhering to the guidelines , comprehensive reporting and referencing . Please cite all arguments with the given sources. The information should be structured as follows: an short answer to the query of the use, a detailed body to underline the answer given. Include links to all sources at the end for further reading.

Fig. 2. System Prompt